REC'D 03 NOV 2004
WIPO PCT

PCT/IB04/3274

PA 1238026

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

October 25, 2004

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE UNDER 35 USC 111.

APPLICATION NUMBER: *60/484,201*
FILING DATE: *June 30, 2003*

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH
RULE 17.1(a) OR (b)

By Authority of the
COMMISSIONER OF PATENTS AND TRADEMARKS

H. L. Jackson

H. L. JACKSON
Certifying Officer

BEST AVAILABLE COPY

Please type a plus sign (+) inside this box → +

modified PTO/SB/16(8-00)

U.S. PTO 60/484201 06/30/2003

PROVISIONAL APPLICATION FOR PATENT COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION under 37 CFR 1.53(c).

Docket Number: 20911-08046

INVENTOR(s)

Given Name (first and middle [if any])	Family Name or Surname	Residence (City And Either State Or Foreign Country)
Takamasa	Koshizen	Saitama, Japan
Bernd	Heisele	Boston, MA
Hiroshi	Tsujino	Saitama, Japan

☐ Additional inventors are being named on separately numbered sheets attached hereto.

TITLE OF THE INVENTION (500 characters max.)

Expectation Maximization of Prefrontal-superior Temporal Network by Indicator Component-based Approach

CORRESPONDENCE ADDRESS

Direct all correspondence to:

☒ Customer Number 00758 →

ENCLOSED APPLICATION PARTS (check all that apply)

☒ Specification *No. of Pages:* 11

☒ Drawing(s) *No. of Sheets:* 4

☐ Application Data Sheet See 37 CFR 1.76

☒ Return Postcard

☐ CD(s), Number

☐ Other *(specify)*

METHOD OF PAYMENT (check all that apply)

☐ Applicant claims small entity status. See 37 CFR 1.27

☒ Fee Transmittal Form Enclosed (in duplicate)

☒ Check Enclosed

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

☒ No.

☐ Yes, the name of the U.S. Government Agency and the Government contract number are: ____________

SIGNATURE OF ATTORNEY OR AGENT

Signature: A. C. Smith

Attorney/Reg. No.: Albert C. Smith, Reg. No. 20,355

Dated: 6/30/03

CERTIFICATE OF MAILING

I hereby certify that this correspondence is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service pursuant to 37 CFR 1.10 in an envelope addressed to: Mail Stop Provisional Patent Application, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450 on the date shown below.

Signature: A. C. Smith

Typed or Printed Name: Albert C. Smith

Dated: 6/30/03

Express Mail Mailing Number: EV 338300100 US

USE ONLY FOR FILING A PROVISIONAL APPLICATION FOR PATENT

PTO/SB/17 (10-02) (Modified)
Approved for use through 10/31/2002. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

FEE TRANSMITTAL for FY 2003

Patent fees are subject to annual revision.

☐ Applicant claims small entity status. See 37 CFR 1.27

TOTAL AMOUNT OF PAYMENT ($) 160

Complete if Known	
Application Number	Not Yet Known
Filing Date	June 30, 2003
First Named Inventor	Takamasa Koshizen
Examiner Name	Not Yet Known
Art Unit	Not Yet Known
Attorney Docket No.	20911-08046

METHOD OF PAYMENT (check all that apply)

☒ Check ☐ Credit Card ☐ Money Order ☐ Other ☐ None
☐ Deposit Account:

Deposit Account Number: 19-2555
Deposit Account Name: Fenwick & West LLP

The Commissioner is authorized to: *(check all that apply)*
☐ Charge fee(s) indicated below ☐ Credit any overpayments
☒ Charge any additional fee(s) during the pendency of this application
☐ Charge fee(s) indicated below, except for the filing fee to the above-identified deposit account.

FEE CALCULATION

1. BASIC FILING FEE

Large Entity Fee Code	Large Entity Fee ($)	Small Entity Fee Code	Small Entity Fee ($)	Fee Description	Fee Paid
1001	750	2001	375	Utility filing fee	
1002	330	2002	165	Design filing fee	
1003	520	2003	260	Plant filing fee	
1004	750	2004	375	Reissue filing fee	
1005	160	2005	80	Provisional filing fee	160

SUBTOTAL (1) ($) 160

2. EXTRA CLAIM FEES FOR UTILITY AND REISSUE

		Extra Claims		Fee from below		Fee Paid
Total Claims	-20** =		X		=	
Independent Claims	-3** =		X		=	
Multiple Dependent					=	

Large Entity Fee Code	Large Entity Fee ($)	Small Entity Fee Code	Small Entity Fee ($)	Fee Description
1202	18	2202	9	Claims in excess of 20
1201	84	2201	42	Independent claims in excess of 3
1203	280	2203	140	Multiple dependent claim, if not paid
1204	84	2204	42	**Reissue independent claims over original patent
1205	18	2205	9	**Reissue claims in excess of 20 and over original patent

SUBTOTAL (2) ($) .00

**or number previously paid, if greater; For Reissues, see above

FEE CALCULATION (continued)

3. ADDITIONAL FEES

Large Entity Fee Code	Large Entity Fee ($)	Small Entity Fee Code	Small Entity Fee ($)	Fee Description	Fee Paid
1051	130	2051	65	Surcharge - late filing fee or oath	
1052	50	2052	25	Surcharge - late provisional filing fee or cover sheet	
1053	130	1053	130	Non-English specification	
1812	2,520	1812	2,520	For filing a request for *ex parte* reexamination	
1804	920*	1804	920*	Requesting publication of SIR prior to Examiner action	
1805	1,840*	1805	1,840*	Requesting publication of SIR after Examiner action	
1251	110	2251	55	Extension for reply within first month	
1252	410	2252	205	Extension for reply within second month	
1253	930	2253	465	Extension for reply within third month	
1254	1,450	2254	725	Extension for reply within fourth month	
1255	1,970	2255	985	Extension for reply within fifth month	
1401	320	2401	160	Notice of Appeal	
1402	320	2402	160	Filing a brief in support of an appeal	
1403	280	2403	140	Request for oral hearing	
1451	1,510	1451	1,510	Petition to institute a public use proceeding	
1452	110	2452	55	Petition to revive - unavoidable	
1453	1,300	2453	650	Petition to revive - unintentional	
1501	1,300	2501	650	Utility issue fee (or reissue)	
1502	470	2502	235	Design issue fee	
1503	630	2503	315	Plant issue fee	
1460	130	1460	130	Petitions to the Commissioner	
1807	50	1807	50	Processing fee under 37 CFR 1.17(q)	
1806	180	1806	180	Submission of Information Disclosure Stmt	
8021	40	8021	40	Recording each patent assignment per property (times number of properties)	
1809	750	2809	375	Filing a submission after final rejection (37 CFR 1.129(a))	
1810	750	2810	375	For each additional invention to be examined (37 CFR 1.129(b))	
1801	750	2801	375	Request for Continued Examination (RCE)	
1802	900	1802	900	Request for expedited examination of a design application	

Other fee (specify)______________________

*Reduced by Basic Filing Fee Paid

SUBTOTAL (3) ($) .00

SUBMITTED BY

Name (Print/Type)	Albert C. Smith	Registration No. (Attorney/Agent)	20,355	Complete (if applicable) Telephone (650) 335-7296
Signature	A. C. Smith		Date	6/30/03

1358228
20911/01000/DOCS/1359725.1

(1)

Expectation maximization of prefrontal-superior temporal network by indicator component-based approach

T. Koshizen, Bernd Heisele, Hiroshi Tsujino

Honda Research Institute Co. Ltd.,
8-1 Honcho, Wako-Shi, Saitama-ken, 351-0114 (Japan Office),
145 Tremont Ave., Boston, MA., 02111-1208 (Americas Office)
koshiz@jp.honda-ri.com

Abstract

In CNS*02, we hypothetically provided the cross-supramodal integration system, which is hypothetically dedicated by a bidirectional neuronal network where prefrontal cortex (PFC) is interacted with hippocampus (HC) in order to calculate the coherent relationship between the supramodalities. The parameter to learn the coherence may also be dedicated by Anterior Cingulate (AC) cortex. In this paper, we attempt to propose top-down attention control system based on the outcome that we presented at CNS*02. That is, PFC is presumably interacted with superior temporal (ST) neuron in order to extract indicator component representing the whole view of face or object, by maximizing the expectant value where attention modulation can be taken into account of distinguishing different faces. As a result, we postulate an objective of the top-down attention control is essentially to compute the abstraction of face/object information based on the Expectation Maximization (EM) algorithm since the voluntary movements of facial viewpoints must play an important role of integrating spatial and temporal property.

Key words: Prefrontal and Superior-temporal Cortex, Distributed Cortical Neuronal Network, Spatiotemporal Attention, Cross Supra-modality, EM Algorithm

1 Introduction

PF cortex is involved in a broad array of cognitive functions, including learning, memory, attention, executive function, planning, and judgment [1]. It is known that PFC has also the executive committee by consolidating hippocampus and other cortical areas. In CNS*01, we suggested the computational

model of anterior-cingulate (AC)-prefrontal-posterior parietal (PP) cortex in relation to attention demanding and modulation process [2]. Conclusively, this work has indicated the aim of PFC-based executive neuronal network was to compute the spatiotemporal attention mechanism. Furthermore, we in CNS*02, presented the computational model of spatiotemporal attention to learn the coherent relationship across the cross-supramodal integration [3]. From recent neurobiological results, it is indicated that the rewiring mechanism can be induced by the inhibitory neuronal network onto excitatory neurons of cerebral cortex where PF neuron may be crucial role of representing the supramodal density and the cross-supramodal correlation by interacting AC cortex. In this paper, we propose the top-down attention control scheme that is hypothesized the computation with respect to the connection of PF and superior temporal (ST) cortex. ST cortex is presumably known a conjunctive point between the 'dorsal' stream specialized by motion (temporal attention) property and the 'ventral' stream (what) specialized by form (spatial attention) property. Additionally, the ST is known for the multimodal response as well as the cross-modal response when interacting with posterior parietal (PP) cortex. Importantly, recent physiological results also demonstrated the neurons in ST involve in the computation of face perception [4]. This is because the two core areas of face processing engage in the categorization of a stimulus as a face, and the identification of a specific individual, by the ST neuron incorporated with inferior temporal (IT) neuron, in order to implement the facial perception consisting of the 'rough' and the 'finer' computations through their neurons belonging in PF and Orbito-Frontal (OF) cortex. The PF-based executive neuronal network involving ST cortex is hypothetically represented as shown in Fig.1, and AC, Basal Ganglia (BG) and Hippocampus (HP) are taken part in the network. The key idea of our speculation is, PF cortex, which may expertise the top-down attention control where attention modulation allows the maximum expectant to yield the semantic abstraction representing specific face or object information in accordance with the supramodal computation. In this sense, the computational role of top-down attention may be implicated to calculate the expectation value where the visual motion may be crucial for maximizing it by adding the biasing signals to the visual form. It evokes generalization properties of biological motion perception using a new class of stimuli that were generated by the spatial and temporal characteristic of morphing among different viewpoint patterns. It has been demonstrated using several monkey's experiments shows that the PF neuron involves in the reward-based learning where spatial information becomes more accurate when reward outcome is expected; more accurate representations of spatial information would as a consequence lead to more accurate behavior, e.g., [5]. In this paper, we suggest the top-down attention control scheme based on the PFC-STS network related to a biological Expectation Maximization (EM) learning algorithm to extract the indicator component based on the top-down attention control.

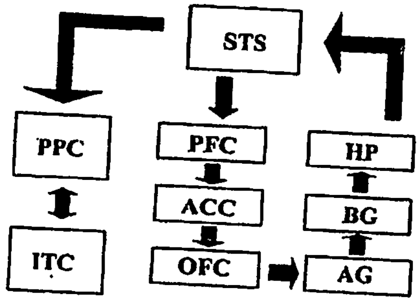


Fig. 1. Hypothetical connection of PFC and STS

2 Computational Model

In this section, we describe the computation model of our proposal, called indicator facial component-based learning approach. The overview of proposed system is shown in Fig.2 (left) where the single component classifier initially developed by [5] is hierarchically reorganized by the multiple component classifiers using Support Vector Machine (SVM) training algorithm which performs pattern recognition for a two-class problem by determining the separating hyperplane that has maximum distance to the closest points of the training set. The closest points, the maximum distance are called Support Vector (SV) and Margin respectively [6]. In our framework, the proposal system aims to extract the indicator facial component that is used to abstract the informative source to distinguish different faces. Face component-based learning algorithm aims to detect faces of different sizes and arbitrary positions in a gray value input image [4]. More precisely, Fig.2. (Right) shows the schematic drawing where component classifiers independently detect components of the face on the first level. This classifier allows the components to extract the features around eyes, nose and mouth. On the second level, the geometrical configuration classifier performs the final face detection by linearly combining the results of the component classifiers. We eventually obtain the output of SVM component classifier indicates if there is a face inside the window or not. Generally, one of the main problems in the component-based object recognition is the selection of the components; how to find discriminated components that allow to distinguish a particular object from rest. To address the question, we employ the proposed system shown in Fig.2 where top-down attention control must be taken into account of maximizing the expectation value where the viewpoint of faces are fluctuated by attention modulation to differentiate the semantic abstraction of indicator component.

Let the expectation value $\mathcal{E}$ be

$$\mathcal{E}(\Omega_\tau) \approx \int logP(\mathcal{O}, \Omega; \tau) d\mathcal{O} \tag{1}$$

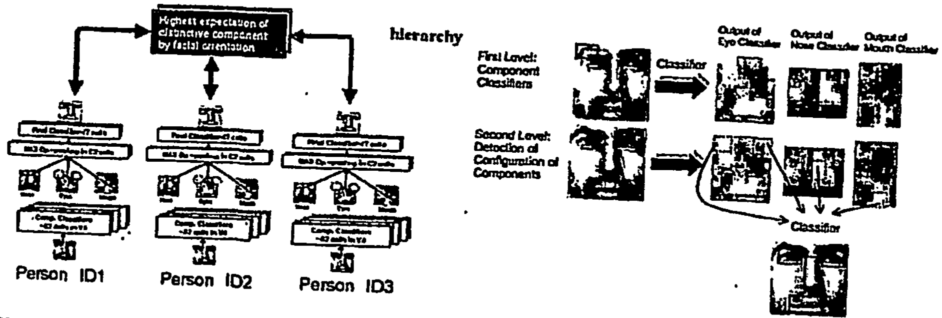


Fig. 2. Proposed multiple classifier system to attain indicator component(Left) and conventional two-level single classifier system (Right)

where Ω represents the probabilistic attention variables *e.g.* see [2] in accordance with certain viewpoints τ. $\mathcal{O}$ is the outputs that are calculated from each SVM component classifier shown in Fig.2 (Left). Additionally, to calculate the probability density function $P(\mathcal{O}, \Omega; \tau)$ we assume the hypothetical mapping $h : \mathcal{O} \mapsto \Omega$. The mapping h will be mathematically defined as,

$$P(\Omega_\tau) \equiv P(y = 1|\mathcal{O}_\tau) = \frac{1}{1 + exp(\mathcal{O}_\tau / \gamma_\tau)} \tag{2}$$

where, y is the classification label of positive examples and $\gamma \in \Omega$ is the predictive parameter of attention modulation to extract the indicator component that maximizes the expectation value $\mathcal{E}$. γ_τ represents the correspondance between the supramodalities such as visual form and motion. Importantly, in our framework the attention Ω can be regarded as the kind of 'hidden variables' calculating to determine the optimal subset of ranked features that is learned from each SVM component classifier. The computation of Eq. (2) is initially suggested by [8]. Mathematically, γ_τ is somehow related to the differentiation of the expectation $\mathcal{E}$ over Ω_τ,

$$\gamma_\tau = \frac{\partial \mathcal{E}(\Omega_\tau)}{\partial \Omega_\tau} \tag{3}$$

Note that the right side of Eq.(3) computes the functional splope of the expectation value $\mathcal{E}$. In principle, we can maximize the logarithm of the joint distribution (which is proportional to the posterior):

$$\Omega_{\tau+1} = argmax_{\Omega_\tau} \;\; \mathcal{E}(\Omega, \Omega_\tau) \tag{4}$$

where τ denotes the parameter representing the viewpoint of a facial component. It is important to remind that the expectation value $\mathcal{E}$ is calculated in the E-step by evaluating the current guess Ω_τ where in the M-step we are optimizing $\mathcal{E}(\Omega, \Omega_\tau)$ with respect to the *free variable* τ (facial viewpoint) to accordingly obtain the new estimate $\Omega_{\tau+1}$.

To implement our indicator component-based approach, training images are captured over the circumstances of various illuminations and the unique (black) background. After the images are collected, pixel values are used as inputs to each layer of a SVM component classifier as shown in Fig.2. The cropped image is then converted into gray values and is re-scaled to 40×40 pixels. Histogram equalization is also applied to remove variations existing in image brightness and contrast. The 1,600 gray values of each face image are then normalized to the range between 0 and 1. Each image is represented by a single feature vector of length 1,600 - the total number of pixels in the image. These feature vectors serve as the inputs to the SVM face classifier during the training process. With respect to the training dataset it includes 974 images of all six subjects in our database. The rotation in depth is again up to about $\pm 42°$. Fig.3 (left) implicates the facial viewpoints has the rotation by right to left, or left to right within $-10°$ to $+10°$. By contrast, the rotation of facial viewpoints is only left by $+12°$ to $+42°$ in Fig.3 (right). Furthermore, Fig.4 shows the expectation values in both cases, and which are calculated based on Eq.(1). Conclusively, their results demonstrate that the expectation value is used for selecting component features by qualifying the input data structure in accordance with different facial viewpoints. We suggest the computation of top-down attention control by PF neuron may be originated to maximize the expectation value where attention class Ω is modulated over different component features, in order to restrict the quantity and quality of training data mapping into the feature space in order to find the optimal subset of selected features, as the indicator component.

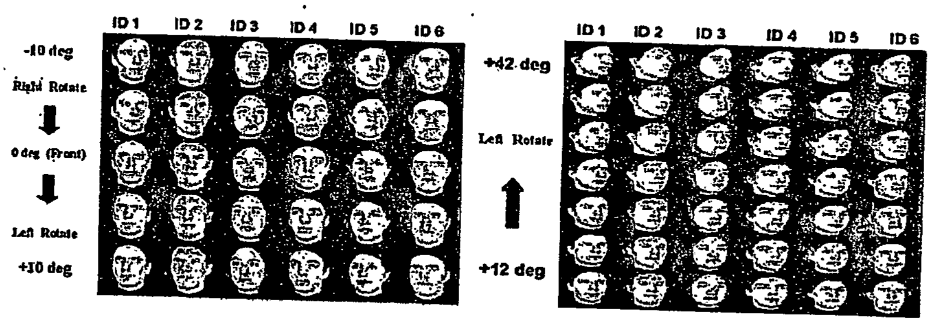


Fig. 3. Facial movement pattern by $-10°$ to $+10°$ (Left) and $+12°$ to $+42°$ (Right)

3 Discussion and Future work

We present the proposed learning system as it relates to component-based face recognition where top-down attention control can be taken into account of facial movements, which are dedicated the possibility that is 'experience-dependent' learning with the intrinsic nature of facial perception. Additionally,

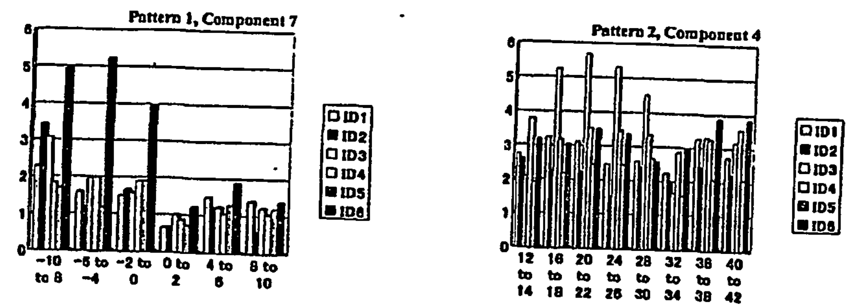


Fig. 4. Histogram shows the learning result by SVM. Vertical axis denotes the margin, while horizontal axis represents the viewpoint discretized by every 2°. ID6 shows the most distinction for the left-side nose component (Left) with −10° to +10°. By contrast, ID3 shows the most distinction for the right-side eye component (Right) with +12° to +42°

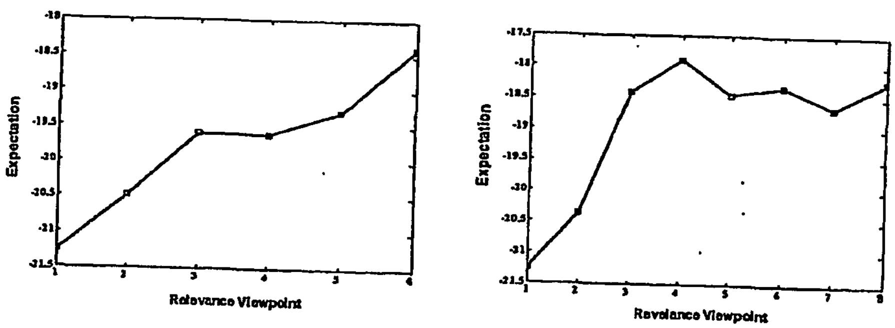


Fig. 5. Expectation value calculated by Eq.1 under the rotation of −10° to +10° (Right) and +12° to +42° (Left)

we assume faces may engender both face-specific and general object processing, and that face-specific processing might be revealed only if the general object system was occupied by concurrent object processing. That is, it is possible that the general object recognition system is itself not *monolithic*, and may become differentiated by experience. It hypothetically allows us to provide the neuronal computation across PFC and STS/IT. Prior to the mechanism of human face recognition, many studies so far indicated that it is the intrinsic

nature itself rather than the experience-dependency. However, several lines of evidence make this conclusion unlikely based on infant experience of face processing [9][10][11][12]. Furthermore, the development of face processing has indicated its significance for social cognition. For example, infants orientate more rapidly to peripheral visual targets when cued by the direction of eye gaze presented by face [13]. Semantic priming refers to the fact that recognition of a word or object of a particular category *e.g.*, animals, is better and faster when preceded by a stimulus of the same category *e.g.*, cat proceeded by dog than when preceded by a stimulus of a different category *e.g.*, cat preceded by pencil. Behavioral studies of face priming indeed have been carried out to infer the processes involved in face recognition. We showed the PFC-STS model suggests the EM algorithm that is basically originated from top-down attention control process biased by visual motion. In our framework, the top-down attention regulates each SVM component classifier by calculating the expectant value to extract indicator component. Future work may allow us the top-down attention to be applying in emotional perception and even in more general-domain for object perception by extending the class of indicator component which is obtained from rotated faces.

References

[1] Miller, E.K. and Cohen, J.D., (2001), *An integrative theory of prefrontal cortex function*, Annual Review Neuroscience, Vol.24, pp.167-202.

[2] Koshizen, T., Akatsuka, K. and Tsujino, H., (2002), *A computational model of attentive visual system induced by cortical neural network*, Neurocomputing, Vol.44-46C, pp.879-885.

[3] Koshizen, T., Yamada, S. and Tsujino, H., (2002), *Semantic rewiring mechanism of neural cross-supramodal integration based on spatial and temporal properties of attention*, Neurocomputing, to appear.

[4] Haxby, J.V. et al., (2000), *The distributed human neural system for face perception*, Trends Cognitive Science, Vol.4, pp.223-233.

[5] Kobayashi, S., Lauwereyns, J. Koizumi, M., Sakagami, M. and Hikosaka, O, (2002), Influence of reward expectation on visuospatial processing in macaque lateral prefrontal cortex, J. of Neurophysiology, Vol.87, pp. 1488-1498.

[6] Heisele, B., Poggio, T. and Pontil, M., (2000), "Face detection in still gray images", A.I. memo 1687, Center for Biological and Computational Learning, MIT, Cambridge, MA.

[7] Vapnik, V., (1998), "Statistical learning theory", John Wiley and Sons, New York.

[8] Wahba, G., (1999), *Support vector machines, reproducing kernel hilbert spaces and the randomized GACV*, In B. Scholkopf, C.J.C. Burges, and A.J.Smola, editors, Advances in Kernel Methods - Support Vector Learning, pages 69-88., Cambridge, MA, 1999. MIT Press.

[9] Valenza, E., Simion, F., Machi Cassia, V., Umilta, C. (1996), *Face preference at birth*, JEP:HPP, Vol. 22, pp.892-903.

[10] Simion, F, Valenza, E., Umilta C. and DallaBarba, B. (1998), *Preferential orienting to faces in newborns: a temporal-nasal asymmetry*, Journal of Exp Psyhol Hum Percept Perform, Vol.24, pp.1399-1405.

[11] Farah M.J., Rabinowitz, C., Quinn, G.E. and Lui, G.T., (2000), *Early commitment of neural substrates for face processing*, Journal of cognitive neuropsychology.

[12] de Schonen S, Mancini, J., Leigeois, F. (1998), *About functional cortical specialization:the development of face recognition*, In the development of sensory, motor and cognitive capacities in early infancy: from perception to cognition, edited by Simon F, Butterworth G., Hove:Psychology Press, pp.103-120.

[13] Farroni, T., Johnson, M.H., Brockbank, M. and Simion, F. (2000), *Infants' use of gaze direction to cue attention:the importance of perceived motion*, Visual cognition, Vol.7., pp.705-718.

Components for Face Recognition
294

Abstract

We present a method for automatically learning a set of discriminatory facial components for face recognition. The algorithm performs an iterative growing of components starting with small initial components located around preselected points in the face. The direction of growing is determined by the gradient of the cross-validation error of the component classifiers. In experiments we analyze how the shape of the components and their discriminatory power changes across different individuals and views.

1. Introduction

In component-based face recognition the classification is based on local components as opposed to the global approach where the whole face pattern is fed into a single classifier. The main idea behind using components is to compensate for pose changes by allowing a flexible geometrical relation between the components in the classification stage. In addition, component-based recognition is more robust against local changes in the image pattern caused by partial occlusion or shadows. In [5] face recognition was performed by independently matching templates of the eyes, the nose and the mouth. A similar approach with an additional alignment stage was proposed in [3]. In [7] recognition was based on Gabor wavelet coefficients that were computed on the nodes of a 2D elastic graph. A comparison between global and component-based face recognition systems using Support Vector Machines (SVM) was published in [1].

The main difficulty in component-based object recognition is the selection of the components, i.e. how to find discriminatory components that allow to distinguish a particular object from other objects. In previous work [2] we introduced an algorithm that learns rectangular facial components for a face detection system. The algorithm starts with small initial components located around preselected points on the face. Each component is grown iteratively; the direction of growing is controlled by an SVM error bound of the component classifier. In the present paper we extend this method to the multi-class problem of face recognition and replace the error bound by the cross-validation error. Not only should the cross-validation error give us a better estimate of the prediction error it also makes the technique applicable to other types of classifiers besides SVMs. For every person in the training database we determine pose-specific sets of components. In our experiments we investigate how the learned components change across individuals and poses. This information might be relevant for a wide variety of face recognition and verification systems.

The outline of the paper is as follows: The image data is described in Section 2. In Section 3 we explain the method for learning components. Section 4 contains the experimental results and the discussion. Section 5 concludes the paper.

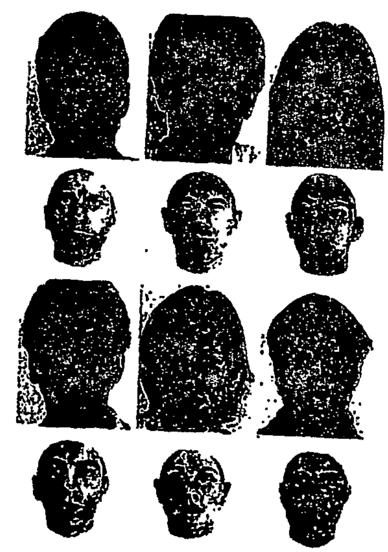


Figure 1: Original images and synthetic images for all six subjects in the database. The face images are arranged according to the ID numbers, starting with person 1 in the upper left corner and ending with person 6 in the lower right corner.

2. Face Data

Our method for learning components requires the extraction of corresponding components from a large number of training images. To be able to automate the extraction process

we used a set of textured 3D head models with known point-wise correspondences. The 3D head models were computed from image triplets (front, half profile, profile view) of six subjects using the morphable model approach described in [4]. Fig. 1 shows an original image and a rendered images of each of the six subjects in the database. Approximately 10,900 synthetic faces were generated at a resolution of 58×58 by rendering the six 3D face models under varying pose and illumination. The faces were rotated in depth from 0° to 44° in 2° increments. They were illuminated by ambient light and a single directional light pointing towards the center of the face. The directional light source was positioned between $-90°$ and $90°$ in azimuth and 0° and 75° in elevation. Its angular position was incremented by 15° in both directions. Example images with different pose and illumination settings are shown in Fig. 2.

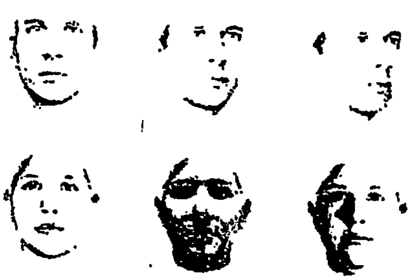


Figure 2: Examples of synthetic faces from the training set. Upper row shows the variations in pose and the bottom row shows different illuminations.

To build the cross-validation set we rendered the 3D head models for slightly different viewpoints and illumination conditions. The faces were rotated in depth from 1° to 45° in 2° steps. The position of the directional light source varied between $-112.5°$ and $97.5°$ in azimuth and between $-22.5°$ and $67.5°$ in elevation. Its angular position was incremented by 30° in both directions. In addition, each face was tilted by $\pm 10°$ and rotated in the image plane by $\pm 5°$. The validation set included about 18,200 face images, some examples are shown in Fig. 3.

Figure 3: Examples of synthetic faces from the validation set. The faces in the cross-validation set were tilted by $\pm 10°$ and rotated in the image plane by $\pm 5°$.

3. Learning Components

An intuitive choice of components for face recognition would be the eyes, the nose and the mouth. However, it is not clear what exactly the size and shape of these components should be and whether there are other components which are equally important for recognition. Furthermore, we would like to quantify the discriminatory power of each component and analyze how the optimal set of components changes over pose and across different subjects. This can be accomplished by an algorithm for learning components which was developed in the context of face detection [2]. The algorithm starts with a small rectangular component located around a preselected point in the face (e.g. center of the left eye). The component is extracted from each face image to build a training set. A component classifier is trained according to the one-vs-all strategy, i.e. the components of one person are trained against the components of all other people in the database. We then estimate the prediction error of each component classifier by cross-validation. To do so we extract the components from all images in the validation set based on the known locations of the reference points. Analogous to the training data, the positive validation set includes the components of one person and the negative set includes the components of all other people. After we determined the recognition rate (CV rate) of the current component classifier on the validation set we enlarge the component by expanding the rectangle by one pixel into one of four directions: up, down, left or right. Again, we generate training data, train an SVM and determine the CV rate. We do this for expansions into all four directions and finally keep the expansion for which the CV rate increases the most. This process can be repeated for a preselected number of expansion steps. We ran our experiments on fourteen components, most of them located in the vicinity of the eyes, nose and mouth (see Fig. 4).

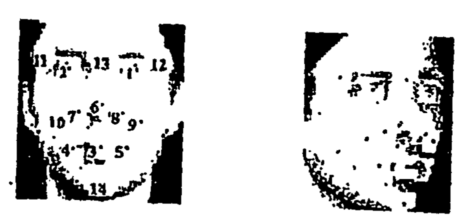


Figure 4: The initial fourteen components for a frontal and rotated face.

4. Experiments

The rotation in depth of the faces in the training and validation sets ranged from 0° to 44°, with increments of 2°. We split the data into three subsets: [0°, 14°], [16°, 30°] and

$[32°, 44°]$, which in the following are referred to as pose intervals 1, 2, and 3, respectively. Each subset included about 630 training and 1060 validation images of each person. To speed-up computation, we randomly removed two thirds of the images in the validation subsets, leaving 350 images in each validation subset. For each person and each pose interval we trained a set of fourteen component classifiers, resulting in 252 component classifiers overall.

In a first experiment we determined the CV rate depending on the width and height of a symmetric component. Symmetric components are rectangular components which are centered on a reference point, i.e. their expansions to the left and right are identical as are the expansions up and down. The dependency on only two variables allows a 3D visualization of the CV rate. We computed the CV rate for SVMs with second-degree polynomial kernel for all sizes of a symmetric component between 5×5 and 21×21 pixels. As features we used the histogram-equalized gray values of the component patterns. Some results are depicted in Fig. 5. The first four rows show the CV rates for the components 2 (right eye), 3 (center of the mouth), and 10 (right cheek) for pose interval 3 (32° to 44°). The last two rows show how the CV rate of component 3 changes across the pose. The surfaces are relatively smooth with few local maxima. It can be expected that gradient-based methods, such as the one described in the previous Section, can be successfully applied to find the maxima. The diagrams also show that the CV rate for a given component strongly changes between subjects, indicating that person-specific components yield better discrimination results than a universal set of components. Another important observation is that the results for a given component/subject combination vary over pose, which suggests the use of view-specific components.

In a second experiment we applied the algorithm described in Section 3 to learn a set of rectangular components. The initial size of our 14 components was set to 9×9 pixels. The number of iterations in the growing process was limited to 10, resulting in 10 components of different sizes and with different CV rates. Of the 10 components we selected the one with the maximum CV rate as our final choice. As for the previous experiment, we trained SVMs with a second-degree polynomial kernel on the histogram-equalized gray values of the components. Fig. 6 shows the learned components across different subjects and views. The average intensity of the component encodes the CV rate, bright values indicate a high CV rate. The pictures show that the CV rates decrease with increasing rotation. This effect is more prominent for components on the left side of the face–the side to which the faces were rotated–than for components on the right side. For the given system, the optimal pose interval for recognizing faces is the near frontal interval, which is similar to results reported in psychophysical experiments on face recognition in humans

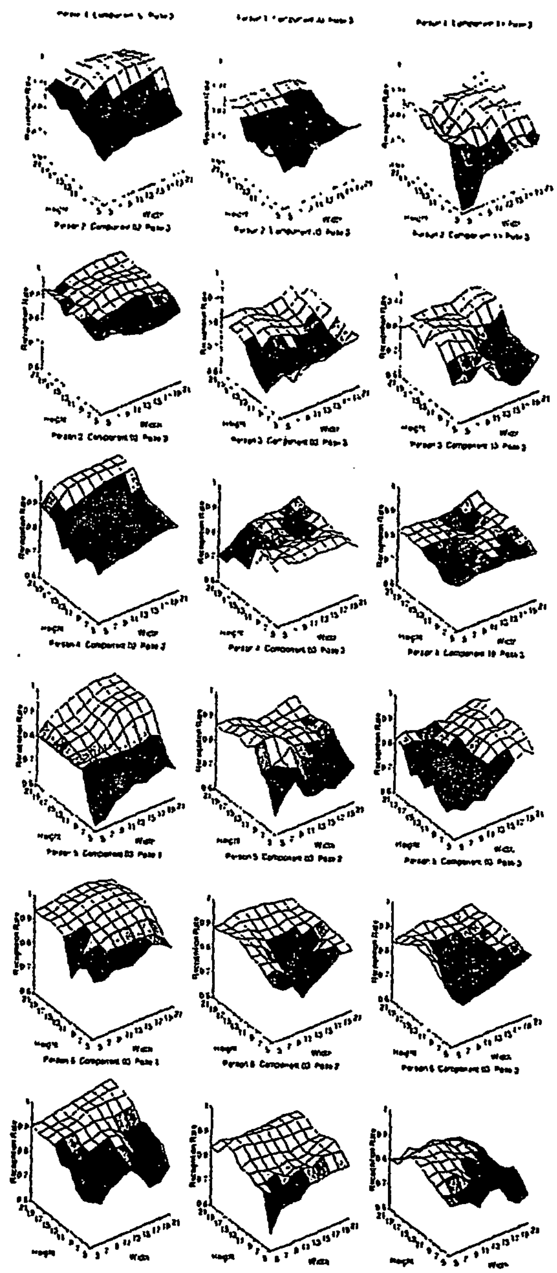
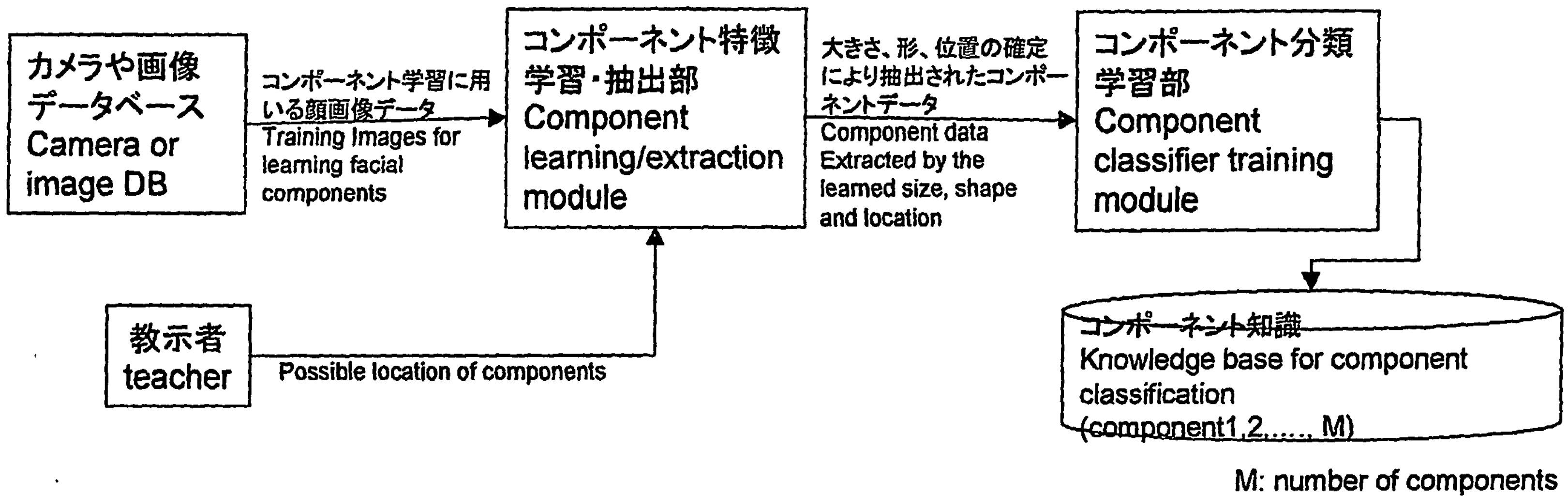


Figure 5: The cross-validation recognition rate for symmetric components. The first four rows show the CV rates for the components 2 (right eye), 3 (center of the mouth), and 10 (right cheek) for pose interval 3: $[32°, 44°]$. The last two rows show how the CV rate changes across the pose intervals $[0°, 14°]$, $[16°, 30°]$ and $[32°, 44°]$.

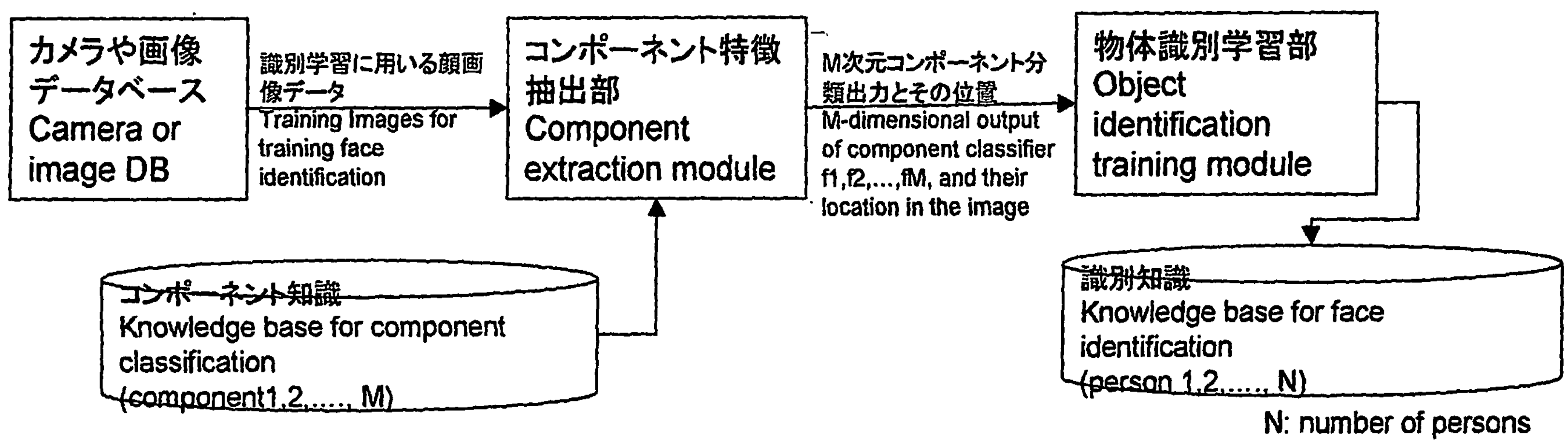
Title: Expectation Maximization of Prefrontal-superior Temporal Network by Indicator Component-based Approach
Applicants: Takamasa Koshizen, et al.
Docket No.: 20911-08046

従来システムにおける学習

1. コンポーネント知識学習フェーズ Component learning phase



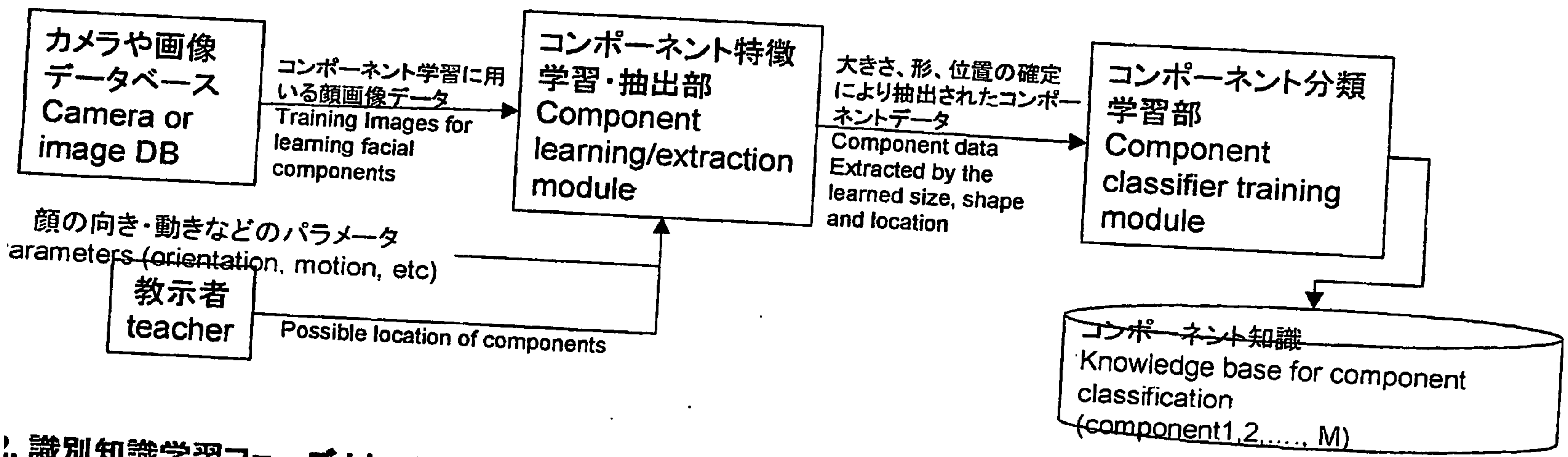
2. 識別知識学習フェーズ Identification learning phase



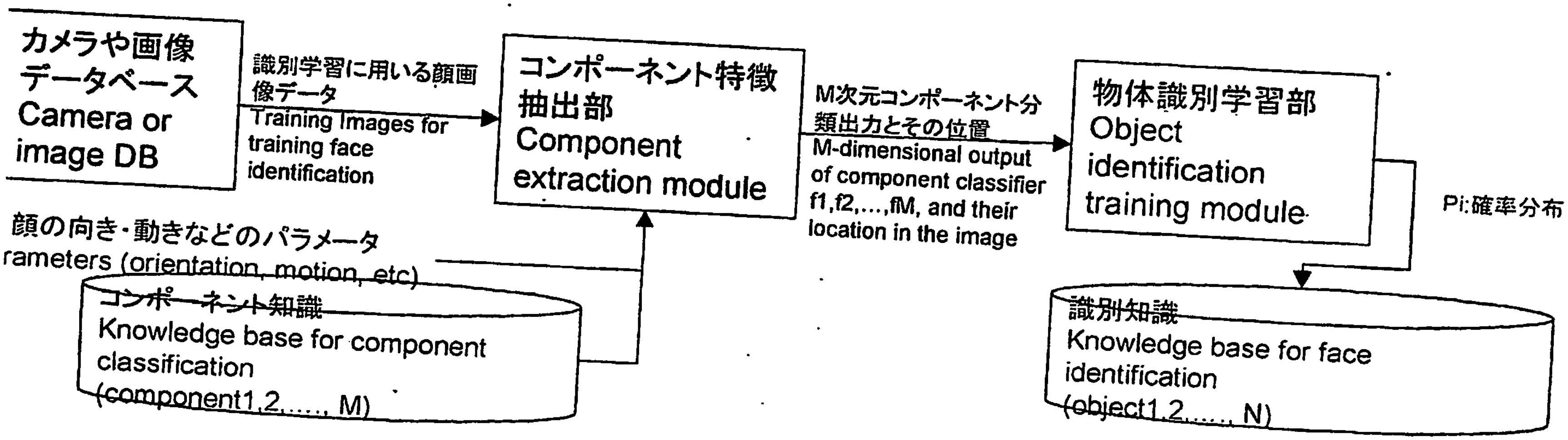
本システムにおける学習

Blue parts are different from the conventional one
Green parts will be different because of the result of modification.

1. コンポーネント知識学習フェーズ Component learning phase



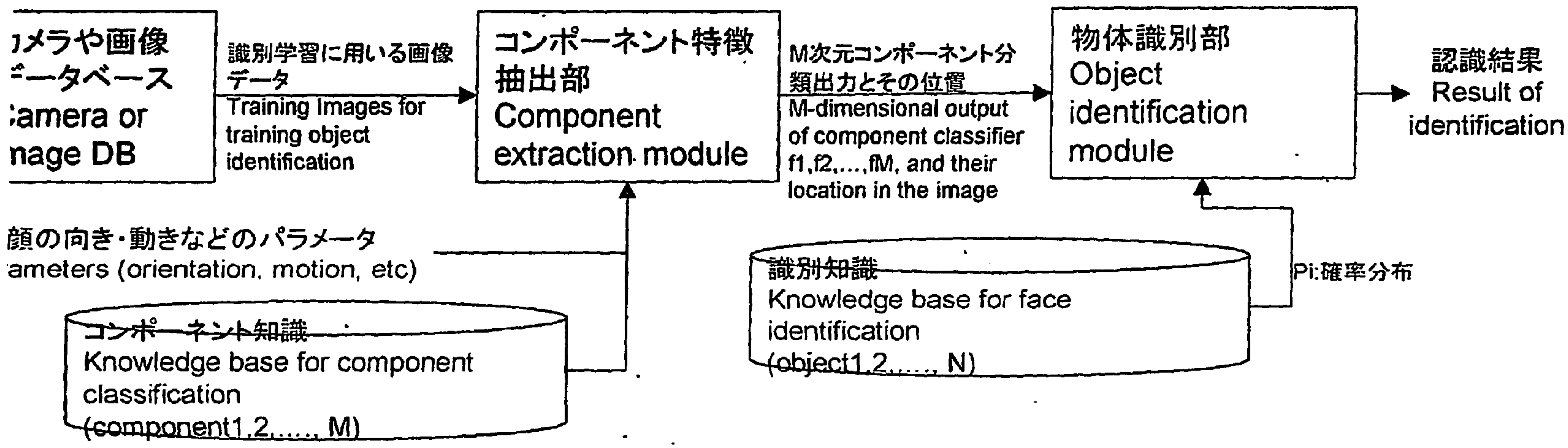
2. 識別知識学習フェーズ Identification learning phase



Title: Expectation Maximization of Prefrontal-superior Temporal Network by Indicator Component-based Approach
Applicants: Takamasa Koshizen, et al.
Docket No.: 20911-08046

システムにおける認識

3. 識別フェーズ Identification phase



4 物体識別学習部について on object identification training module